How Much NMR Data Is Required To Determine a Protein–Ligand Complex Structure?

Ulrich Schieborr, Martin Vogtherr, Bettina Elshorst, Marco Betz, Susanne Grimme, Barbara Pescatore, Thomas Langer, Krishna Saxena, and Harald Schwalbe*[a]

Here we present an NMR-based approach to solving protein–ligand structures. The procedure is guided by biophysical, biochemical, or knowledge-based data. The structures are mainly derived from ligand-induced chemical-shift perturbations (CSP) induced in the resonances of the protein and ligand-detected saturated transfer difference signals between ligands and selectively labeled proteins (SOS-NMR). Accuracy, as judged by comparison with X-ray results, depends on the nature and completeness of the experimental data. An experimental protocol is proposed that starts with calculations that make use of readily available chemical-shift perturbations as experimental constraints. If necessary, more sophisticated experimental results have to be added to improve the accuracy of the protein–ligand complex structure. The criteria for evaluation and selection of meaningful complex structures are discussed. These are exemplified for three complexes, and we show that the approach bridges the gap between theoretical docking approaches and complex NMR schemes for determining protein–ligand complexes; especially for relatively weak binders that do not lead to intermolecular NOEs.

Introduction

Knowing in great detail the molecular structures that are responsible for the interaction of a protein with a ligand is of utmost importance for drug design.[1] Due to our increasing understanding of intermolecular interactions, much effort has been put into designing novel ligands in silico, once a protein structure of a related complex has been solved by X-ray crystallography or NMR.[2] In most real cases, however, a purely theoretical approach is not feasible because the simulations have not reached the level of sophistication that would allow the prediction of binding sites or account for complex three-dimensional rearrangements. Thus, there is a need for fast and reliable experimental solutions of 3D protein–ligand complex structures.

Protein–ligand structures are now routinely obtained by X-ray analysis of ligand-soaked protein crystals.[3] Nevertheless, there are numerous cases in which this approach fails due to the problems of soaking the crystals with certain ligands, multiple binding modes, or large-scale motions of the protein that disrupt the crystal lattice. Alternatively, NMR spectroscopy can, in principle, provide intermolecular distance restraints between protein and ligand that are sufficient to solve a protein–ligand structure. The use of isotope-labeled protein and unlabeled ligand, which allows unambiguous detection of NOEs, has been particularly useful in this respect.[4] Complex structures require ${}^{13}C$-labeled protein, a near-complete protein assignment including side chains, and complicated experimental schemes. Therefore, NMR has a much lower throughput than crystal soaking. Even worse, it requires moderately tight binders. For both reasons, NMR is often not the method of choice for solving ligand–protein complexes, particularly in the early stages of drug development when usually large numbers of binders with weak binding affinities have to be structurally characterized. This is even more true for recent approaches that search for "fragments"[5] that are to be tethered together once their individual binding modes are known.

Several straightforward and faster approaches for overcoming this problem have been presented. Selective-labeling approaches have been used to measure intermolecular NOEs without the need for explicit side-chain assignment.[6] Ligand-based techniques, such as TrNOE[7] or TrCCR,[8] define the structure of a ligand in its protein-bound state but do not give information about its binding mode. Sometimes they even require isotope labeling of the ligand. The "structure–activity relationships by NMR" (SAR by NMR) method,[9] makes use of ligand-induced chemical shift perturbations (CSPs) in the protein to localize binding sites. This method requires only a backbone assignment and a 3D structure (which can also be a homology-modeled structure) of the protein.

It is evident that CSPs can also guide modeling. For example they can define the orientation of a ligand relative to the protein.[10] It has been shown that it is possible to generate meaningful structures from CSPs in protein–protein interactions.[11]

Here, we describe how this approach can be successfully extended to calculate complexes of small ligands bound to pro-

[a] Dr. U. Schieborr, Dr. M. Vogtherr, Dr. B. Elshorst, Dr. M. Betz, Dr. S. Grimme, Dr. B. Pescatore, Dr. T. Langer, Dr. K. Saxena, Prof. Dr. H. Schwalbe
Institute for Organic Chemistry and Chemical Biology
Center for Biomolecular Magnetic Resonance
Johann Wolfgang Goethe Universität Frankfurt
Marie Curie Straße 11, 60439 Frankfurt am Main (Germany)
Fax: (+49) 69-798-29515
E-mail: schwalbe@nmr.uni-frankfurt.de

teins in relatively high throughput. We concentrate on three main questions: i) what are the selection criteria that identify an accurate structure? ii) how many NMR data are actually needed for the calculation of a structure?, and iii) how accurate is this structure and how does it depend on the nature of experimental input?

Results and Discussion

The ambiguous interaction restraint (AIR) concept in protein–ligand interactions

In the traditional concept of NMR structure determination, distance restraints from NMR data are usually introduced between a defined pair of atoms. They lead to a "penalty" energy if a target distance is violated. By contrast the AIR approach, which is used in protein–protein docking (HADDOCK)[11] and was first introduced into automated structure calculation/assignment process of proteins by Nilges and co-workers,[12, 13] assumes a multitude of restraints that are applied between two sets of atoms. The penalty energy calculated from these is defined so that the fulfillment of a restraint causes a steep decrease in the energy function. However, a distance increase in a clearly unfulfilled restraint causes only a smooth increase in the energy term. The aim of the simulation is then to find reasonable physical structures with a maximum number of fulfilled AIRs.

CSPs as a source for AIRs were first proposed by Nilges et al.[13] in 1998 and are increasingly utilized for the determination of protein–protein complexes.[11] Usually, AIRs are defined between all atoms of two residues that experience CSPs. This is not reasonable for protein–ligand interactions because the ligand is much smaller than the protein. If all atoms in a perturbed protein residue were used in AIR definition, the calculation would achieve an optimal solution solely by side-chain reorientations on the protein, and the full range of ligand orientations would not be sampled. Therefore, only the amide protons of shifting residues were used in AIR definition. Thus, all AIRs are defined only in one direction, from the ligand to the protein. This means that every CSP leads to one AIR.
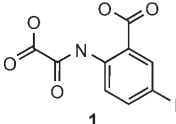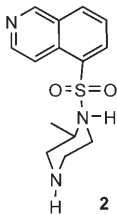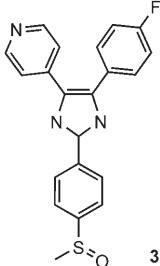
Selection of structures

Structure determination by NMR yields an ensemble instead of one unique configuration. This is because the structure is underdetermined with respect to the underlying data. The deviation of any structure from the "real" one, which is principally unknown, is called the accuracy of the structure. Defining criteria for choosing structures with good accuracy has been an essential part of NMR-structure calculation validation.[14] A "reference" crystal structure is used to calculate the accuracy, which in our case is simply the root-mean-square (rms) difference in the coordinates of the ligand, between the crystal structure and the outcome of the NMR calculation. With this approach, it turned out that correct protein structures could be obtained, in general, with an appropriate energy function during the molecular-dynamics (MD) simulation.

Intermolecular forces that are of particular importance for protein–ligand complexes are difficult to quantify. Therefore it is difficult to make assertions concerning the accuracy of ligand orientation, although significant process has been made in the field.[15, 16] On the other hand, we found that the energy hyper-surface that is defined by AIR is too simple to be an independent measure of accuracy. Therefore, a suitable combination of selected experimental and nonexperimental energies is necessary to select structures with good accuracy.

The procedure was tested by using protein–ligand complexes with known X-ray structure and NMR assignment (Table 1). The choice of the starting structure is important for

Table 1. Structures: protein, ligand, target complex structure, starting conformation of the protein, differences between the two primary sequences of the former two structures, and BMRB code of the protein NMR-assignment.

Protein	Ligand	PDB code complex reference	PDB code starting conformation	Mutations	BMRB code
PTP1b	1	1ECV[18]	1PTY[19]	T151S C215S D252E	5474[20]
PKA	2	1YDR[21]	1CMK[22]	M6K Y69F F108Y A124P	6183[23]
p38	3	1A9U[24]	1P38[25]	L48H T26A	6468[26]

the results. In many cases holo structures seem to be the best choice.[17] Nevertheless, calculations were started with coordinates of the apoprotein to prevent the introduction of a "structural bias" by the holo conformation. The calculated complex structures were then compared to the coordinates of the respective complexes. This yields the accuracy that can be correlated to a variety of selection criteria. A principal-component analysis of all energy terms from the simulation with respect to the accuracy showed a very different contribution from the various energy terms. Purely intramolecular terms, such as the bond lengths or angles, are uncorrelated to accuracy. Electrostatic interactions had strong correlations but differed strongly for different proteins. Therefore both are unsuitable as a general measure of accuracy. On the other hand, both the intermo-

lecular van der Waals energy and the experimental AIR energy showed strong correlations of similar size for different protein–ligand systems. They are therefore suitable as measurements of accuracy. Accurate structures are picked from a "selection plot" in which both the intermolecular van der Waals and experimental AIR energy are plotted (Figure 1). Structures with both a low van der Waals and AIR energy are possible solutions. By contrast, structures in which only one of the two energy terms is low are discarded.
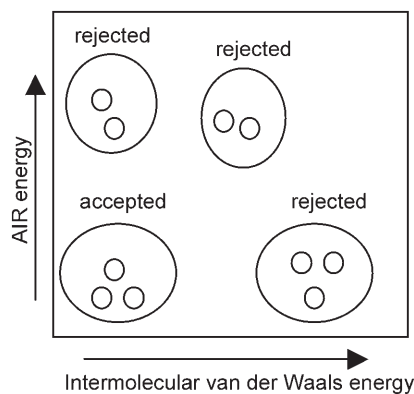

Figure 1. Basic principle of the selection plot. Intermolecular van der Waals energies and experimental AIR energies are evaluated for each complex structure and plotted. Structures that possess both low van der Waals and AIR energies are accepted. Structures are rejected if one (or both) of the two energies is high.

Three case studies highlight the potential and commonly encountered problems of the NMR-based docking approach (Table 1). These are the complexes of i) phosphotyrosine phosphatase 1b (PTP1b) with 5-iodo-2-(oxalylamino)benzoic acid (**1**),[18–20] ii) catalytic subunit cAMP dependent protein kinase A (PKA) with H7 (**2**),[21–23] and iii) mitogen-activated kinase p38 with SB203580 (**3**).[24–26]

Case study 1: PTP1b

CSPs caused by **1** lead to a well-defined binding site (Figure 2a) but few distant residues are also affected. Only residues that could be in the vicinity of the ligand would be transformed to AIRs after determination of the binding area in an explorative structure-determination process. Nevertheless, all experimental CSPs, including the distant ones, were introduced to demonstrate the stability of the calculation. The primary sequences of apo and **1**-complexed PTP1b that have been solved by X-ray crystallography differ in the vicinity of the binding site, as defined by CSPs. Therefore, the apo structure was modified to accommodate the mutations T151S, C215S, and D252E that are present in the complex. Distant AIRs cause severe van der Waals violations in the rigid body minimization step. Therefore, AIRs were switched off in all subsequent calculations. Experimental parameters are thus switched on at the initial stage of the procedure and help in finding the binding site. All later steps are purely modeled optimizations of the protein–ligand interactions. The resulting selection plot has an
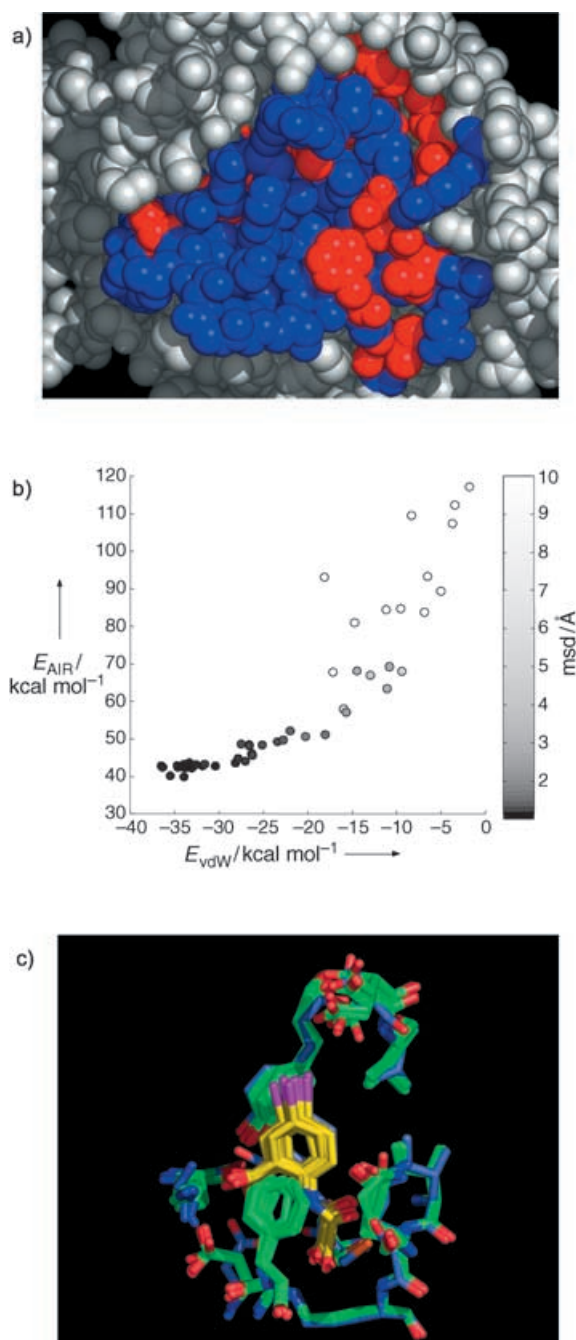

Figure 2. a) Residues with the largest CSPs (red) caused by **1** and residues for which flexibility was allowed during the simulation (blue), mapped onto the X-ray structure of apo-PTP1b. b) Selection plot of the resulting complexes. The accuracy of each structure relative to the reference X-ray complex is indicated in gray scale. This represents the positional rmsd in the coordinates of the ligand. The selection plot shows that in this case one of the two energies would be sufficient for the selection of a structure. c) A selection of accepted structures from NMR data (green/yellow) corresponds well with the X-ray reference complex (blue), with an rmsd of ca. 1 Å.

excellent correlation of intermolecular van der Waals and AIR energies with the positional root-mean-square deviations (rmsd) of the ligand, relative to the reference structure (Figure 2b). Complexes with low energy for both energy terms have low rmsd relative to the reference structure. Equally im-

portant, both energies rise dramatically if the structure deviates from the reference structure. This is particularly interesting for the artificial AIR energy since this energy was switched off during the final simulation steps. The result of this structure calculation is a complex conformation which has a positional rmsd relative to the X-ray structure between 1.1 Å and 1.5 Å in the coordinates of the ligand (Figure 2c).

This case study shows that the calculation of a complex structure from CSPs can be straightforward. No further assumptions concerning known binding modes were used as input. Furthermore, AIRs far away from the binding site do not bias the final result, provided that they are switched off during the refinement process. It is also interesting to note that each of the energies alone (experimental or van der Waals energies) would have been sufficient to select the correct structures.

Case study 2: PKA

For PKA, CSPs are located in a well-defined cluster around the adenine-binding pocket. Therefore, all AIRs remained switched on during the entire simulation. Since any sequence differences of the various PKA structures are more than 15 Å away from this site, neither of these mutations had to be considered in the simulation.

Although the CSPs seem to define the binding site reasonably well, it turns out that the basis of available NMR and X-ray data complicates the situation; this is very common in drug discovery. First, the NMR assignment[23] is incomplete and excludes a large part of the adenosine-binding pocket surface. Second, although the apo structure is known (PDB code 1J3H)[27] there is no electron density for large (though not essential) parts of the protein. The structure of protein kinase inhibitor (PKI) bound PKA (PDB code 1CMK),[19] which was used as the starting model, deviates with 1.2 Å backbone rmsd from apo-PKA, and the PKA–**2** complex (PDB code 1YDR)[21] deviates by another 1.8 Å. This is due to the conformational switch that changes PKA from its "open" apo conformation to a "closed" ligand-bound conformation. These forms are only extremes of an ensemble of conformational states that have been characterized for PKA; the choice of the appropriate starting structure is therefore problematic. Notably, a related equilibrium between open and closed conformations also exists for the WPD loop in PTP1b.[28] However, the selection of the starting structure is easier since 1) the conformational change is only local, and 2) it manifests itself in strong CSPs of the loop residues. For PKA, both structures were tentatively used as starting structures.

Starting from the protein in a closed form (PDB code 1ATP) fails in the initial rigid body minimization step, simply because the closed conformation prevents the ligand from entering its correct binding site. It is needless to mention that the simulation is very successful if the ligand is placed initially at the binding site (data not shown). This demonstrates that the proper choice of a starting protein structure is essential. It was previously noted that the outcome of virtual screening was largely dependent on the target protein structure.[15]

When starting from the apo structure (PDB code 1CMK), which is an open conformation, the resulting configurations after water refinement are not closer than 3 Å rmsd from the reference structure in the coordinates of the ligand. Based on the selection plot the identification of meaningful structures is nevertheless possible (Figure 3b). The difference is at least partially due to an incorrect positioning of the piperazine ring. This moiety has little shape complementarity and few interactions with the protein. Hence it can be modified to methylaminoethyl in H8, which binds only twice as tightly as H7. On the other hand, the orientation and possible constructive interactions of the quinazoline ring, which is the main feature of the H series of inhibitors,[21] are correctly reproduced. Both the X-ray and lowest AIR-energy structures show the hydrogen bridge to the backbone of Val123. This information can thus guide possible drug development programs. This is an encouraging result bearing in mind that i) the simulation does not in any way reproduce the drastic conformational changes of the protein, particularly the open/close transition, and ii) considerable parts of the ligand-binding site of the protein are unassigned and do not contribute to the experimental input. This example shows that the combination of two selection criteria is essential (Figure 3b) since both favorable van der Waals interaction and experimental energies are also compatible with high rmsd deviations.

This case study also shows that even for a protein with incomplete information a complex structure can be calculated from the CSP data alone. Although neither CSP data nor the calculation predict the changes in the domain orientation, the binding mode is correctly elucidated.

Case study 3: p38

p38 is similar to PKA in that the ligand-induced CSPs cluster in a well-defined region that is far away from any differences in the primary sequences of apo and reference structures. The definition of the binding site (Figure 4a) is more complete compared to PKA because the extent of assignment is higher.[26] Since apo- and ligand-bound structures deviate by only 0.9 Å, the choice of a proper starting structure is less critical than in the case of PKA.

Despite this promising situation, the calculation yields a multitude of solutions with comparable AIR and van der Waals energies (Figure 4b). The correct structure is among these but it is not possible to identify it as the solution of the docking problem. The reason is the specific shape of SB203580 (**3**), which has one twofold and one threefold axis of rotational symmetry. This implies that the ligand can also occupy the binding site in other symmetry-related orientations. This failure does not affect the principle of the selection plot. Since different binding modes lead to very similar final energies, neighbored points in the selection plot have a high rms difference. By contrast, neighbored structures in the selection plot of the two examples discussed above belong to the same structural family. Whether an ensemble of configurations from the selection plot is a proper choice or not is thus readily decided from their relative structural differences.
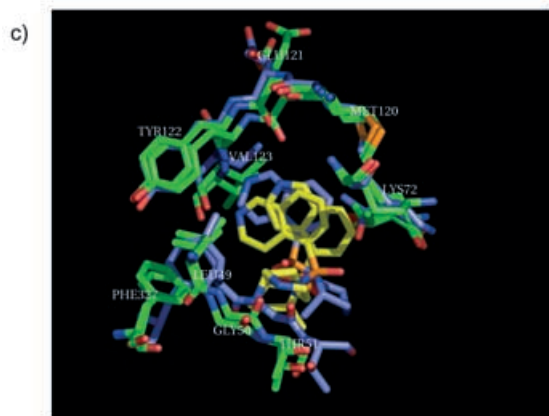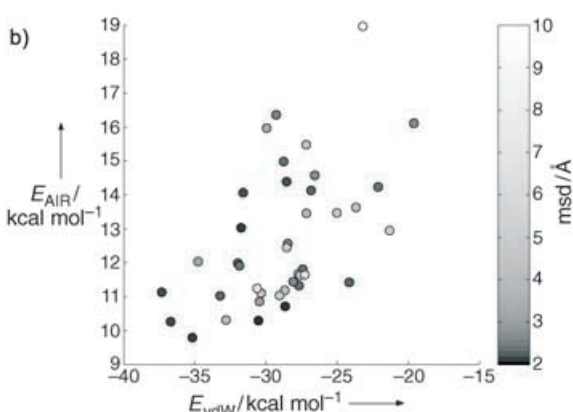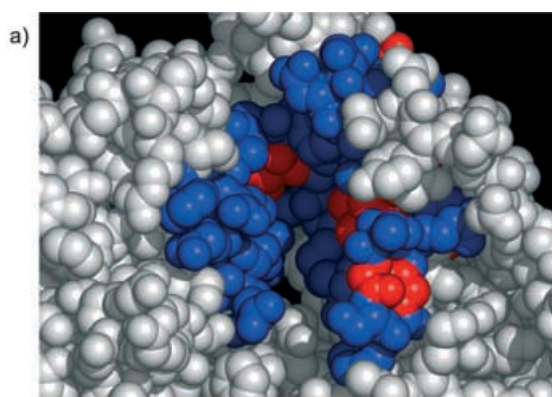
Figure 4. a) Residues with the largest CSPs (red) caused by **3** and flexible residues (blue) mapped onto the X-ray structure of apo-p38. b) Selection plot starting only from CSP data. The accuracy of each complex relative to the reference X-ray structure is indicated as described for Figure 2. The calculation leads to many accurate and wrong structures with comparable energies. It would be impossible to pick a single structure out of these.

Figure 3. a) Residues with the largest CSPs (red) caused by **2** and flexible residues (blue) mapped onto the X-ray structure of apo-PKA. b) Selection plot of the resulting complexes as described in Figure 2b. The accepted structures are not the ones with the best individual energies. c) The two best complexes from NMR data (green/yellow), compared to the X-ray reference structure (blue), show good agreement of the quinazoline ring. This ring is responsible for the typical interactions observed for the H class of PKA inhibitors. The agreement of the nonconserved piperazine ring is less good, and results in an overall rmsd of 3 Å relative to the reference X-ray structure.

Whereas it is straightforward to resolve this ambiguity by X-ray crystallography due to different diffraction properties of the three rings, this is not possible by CSPs which only define a binding site. In order to yield an unambiguous solution, the structure calculation needs additional input. This can be either restraints that enforce plausible interactions or additional NMR
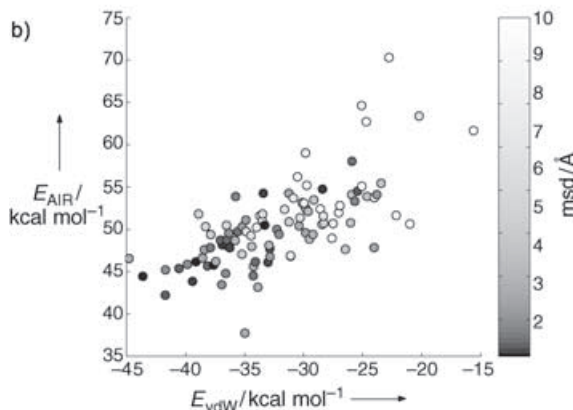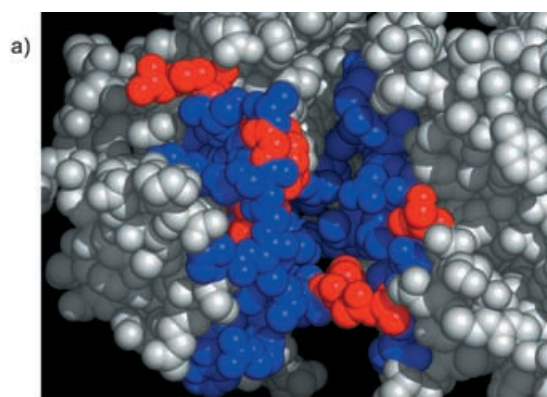
data. We discuss both options since both were found to be successful.

The first approach, which restrains the simulation to a plausible solution, is based on the wealth of structural data that is available for protein–ligand complexes in general and for protein kinases in particular. In the case of the p38–**3** interaction CSPs localize the binding at the adenosine-binding site. In nearly all known cases binding at this site includes formation of at least one specific hydrogen bond to the "hinge" residues (His107–Met109 in p38). These manifest themselves in very strong CSPs for these amino acids, and are in principle measurable by NMR since they influence exchange properties with bulk water. It is reasonable to restrain the simulation such that at least one hydrogen bond with one of the hinge-residue amide groups is formed. We have done this by introducing one additional AIR between possible ligand and the hinge donor/acceptor pairs. The definition of this single AIR is sufficient to make the result of the simulation unambiguous (Figure 5a). This approach shows that apparently the hydrogen bonds are not correctly modeled within our simulation and supports the notion that the modeling of protein ligand-binding sites is challenging[29] and still insufficiently reproduced in common force fields.
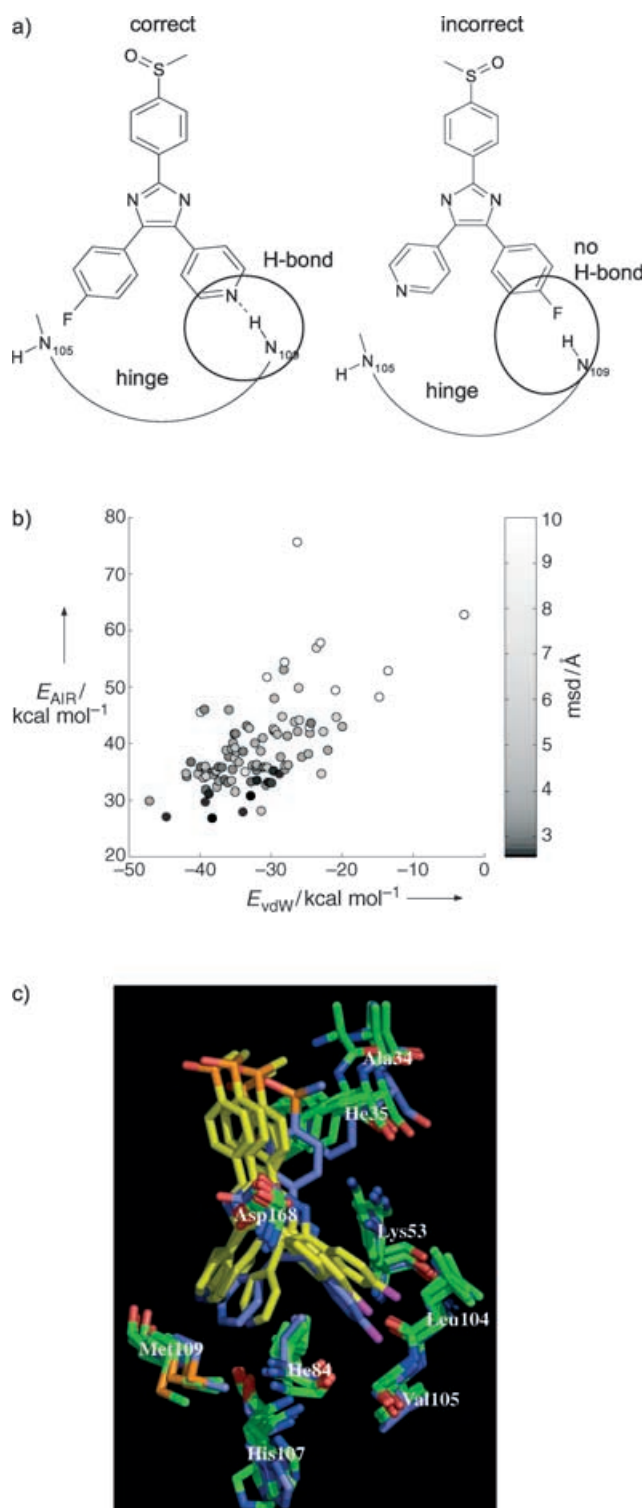
Figure 5. The high symmetry of SB203580 (**3**) causes several problems. Particularly the C_{2v} symmetry of the two rings is difficult to resolve. a) The two possible options that differ in their ability to form a hydrogen bond with the Met109 amide group. These two options highlight the difficulty in quantifying hydrogen bonds in molecular simulations. b) Selection plot of the complexes resulting from a simulation that includes both experimental AIRs and the assumption of at least one hydrogen bond. c) A selection of accepted complexes from NMR data (green/yellow) compared to the X-ray reference structure (blue) shows that the NMR-derived docking mode agrees well with the reference X-ray structure (rmsd of 1.8 Å–2.9 Å in the coordinates of the ligand).

The second approach to resolve the ambiguity of the ligand is the inclusion of additional experimental NMR data. Whereas differential line broadening or saturated transfer difference (STD) buildup of ligand signals are inexpensive and quick methods, their information content is not always high enough to unambiguously characterize the ligand-interaction surface. More sophisticated recent approaches aim at a quantification of STD buildup, either by back-calculation of the full relaxation matrix (Corcema-STD)[30] or by simplification of the relaxation matrix by selective labeling (SOS-NMR).[14]

As an example, we show how the SOS-NMR approach yields a unique solution of the problem. The SOS concept was originally demonstrated with selectively [1]H-labeled and otherwise deuterated proteins. The selectively labeled amino acids lead to patches of STD-active surface that cumulatively yield a unique solution. Every single sample has a high degree of ambiguity that is reduced stepwise by the overlap of a sufficient number of other samples. In combination with CSPs, the number of required SOS-samples is smaller because CSPs define the binding site accurately enough to enable the search for residues that occur only once near this binding region. We restricted our experiments to amino acids that: i) can be over-expressed in deficient *E. coli* strains, to ease the labeling process, ii) bear methyl groups, to simplify the STD experiment, and iii) are less abundant in the ligand binding site (as defined by the CSPs) to simplify the evaluation. An analysis of the p38 X-ray structure shows that Ile84 is present near the binding site and can lead to STD effects (Figure 6a). Experimental data (Figure 6b) essentially confirm what is expected from the geometrical analysis. An additional AIR connecting the side chain protons of Ile84 and protons of the fluorophenyl and pyridyl rings was constructed.

The introduction of only a single restraint from the SOS sample does not lead to meaningful results but yields the correct solution, among others. This is not unexpected because distance restraints involving just one methyl group cannot uniquely define a three-dimensional structure. Also, spin diffusion effects that are expected for a protein of this size are more likely to falsify the distance information of NOEs than in the published example of the 12 kDa FKBP (FK506-binding proteins).[14] However, if the SOS results are combined with the CSP data, this leads to a unique solution (Figure 6c). Obviously the proper choice of the labeled amino acid (or the use of more than one sample) is essential for the outcome of the experiment.

In addition to the methods exploited here, particularly the back-calculation of ligand CSPs is a very promising approach.[31]

Conclusion

The aim of this study was the calculation of protein–ligand complexes that are obtained from a minimal set of NMR-derived information. It turns out that surprisingly little NMR information is necessary to yield a unique complex structure. In the search for selection criteria that identify an accurate family of structures we introduced a selection plot. It utilizes the fact
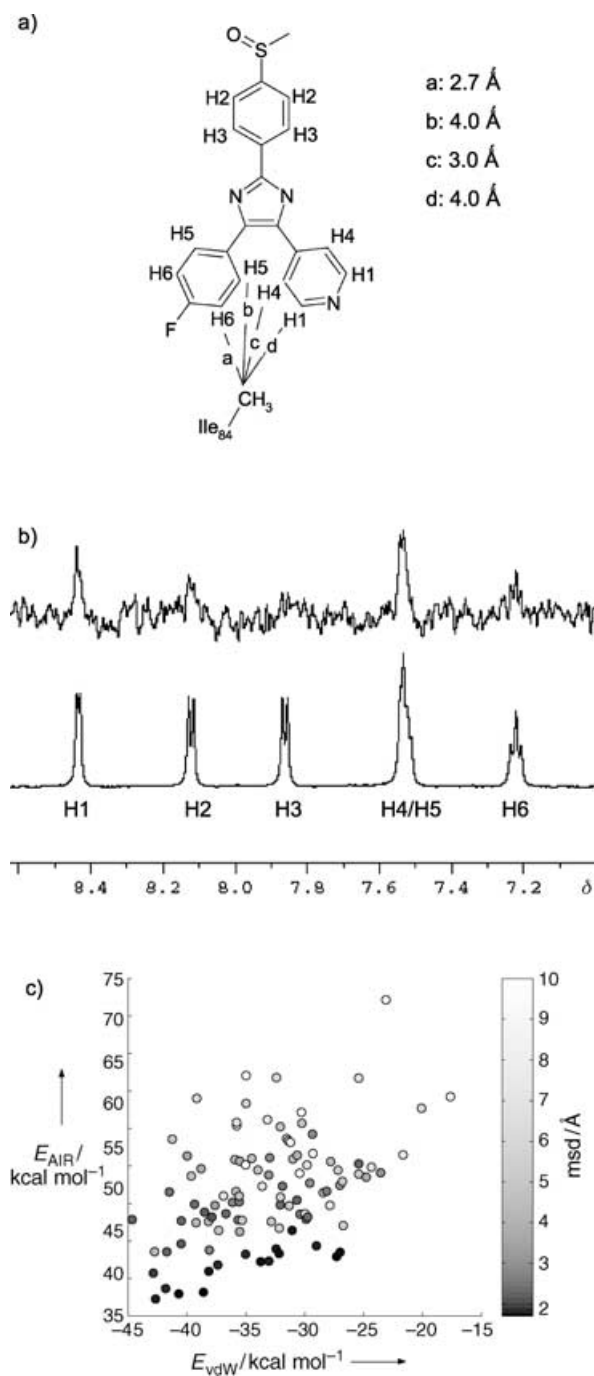
Figure 6. a) The use of the SOS-NMR principle with deuterated and selectively protonated protein can resolve the orientation of the ligand. These samples lead to STD effects at short saturation times only for ligand protons that are closer than 6 Å to the protonated site. The nearest isoleucine, Ile84, leads to saturation transfer at both rings and can orient the ligand correctly. b) The experimental SOS-NMR data confirms the assumptions that are expected from the geometric analysis of the X-ray structure. c) Selection plot of the complexes resulting from a simulation that includes CSP and SOS data. In both cases an unambiguous solution is found.

that the combination of the intermolecular van der Waals energy from the physical force field and the penalty energy from the violation of the experimental parameters is sufficient to identify a correct structure.

The structures presented in this study are derived from very sparse NMR data. Although the resulting structural models are not very accurate if compared to a high-resolution NMR structure, the essential interaction partners and the orientation of the ligand, which are of prime importance for the medicinal chemist, are correctly reproduced. Obviously, the identification of a binding area by CSPs suffices in many cases to identify the correct orientations of the ligand with respect to the protein. It seems to be more important than complete NMR assignments and CSP interaction profiles (which are very often not available for typical targets), or even the choice of a proper starting structure. Once a binding site is given, only a limited number of orientations can satisfy both the surface complementarity of protein and ligand and constructive interactions. It has been shown that any protein–ligand interaction can occur at such "hot spots" that possess particular surface and shape properties.[29] Although this result can, in principle, also be obtained from purely theoretical modeling, NMR is indispensable for the identification of protein hot spots. It is intriguing that the only real problem, the docking of SB203580 (**3**) to p38, is caused by the high symmetry of the ligand. We could show that the inclusion of a few additional experience- or experiment-based restraints suffices to solve the symmetry problems. Due to the "open" concept of "crystallography and NMR system" (CNS),[32] many other restraints are conceivable.

NMR structure determination of protein–ligand complexes has come a long way, even though in its stringent classical form, it is not a real alternative to X-ray structure determination. On the other hand, high throughput crystallization programs[3] readily identify binding and binding mode at the same time. NMR has been successfully applied to the screening mixtures and finding binding sites.[9] Here we show that in many cases the additional step to binding modes is only marginal, a procedure which we would like to call LIGDOCK.

Experimental Section

Origin of CSPs: The procedure proposed in this article was validated by application to three protein–ligand complexes with known X-ray coordinates for apoproteins and protein–ligand complexes and known NMR assignments as compiled in Table 1. Differences in the primary sequences of the apo structures, the complex structure, and the construct used for NMR measurements were treated as mentioned in the results section.

Proteins (PTP1b$_{1-282}$, PKA$_{1-350}$, p38$_{2-349}$) were overexpressed with uniform ^{15}N-labeling in *E. coli* following standard procedures. All CSPs were extracted from 800 MHz TROSY spectra by using published assignments (Table 1) as chemical shift index ($\Delta_{1H+15N} = (\Delta^2_{1H} + 0.17 \cdot \Delta^2_{15N})^{1/2}$). CSPs were classified as strong, medium, and weak based on the shift distributions, and only strong CSPs were used for the calculation of structures (Table 2). All experimental restraints were applied as AIRs. The binding area of the protein can be located in all three examples by mapping the strong CSPs onto the protein structure.

MD simulations and analysis of resulting structures: A three-stage docking protocol, which was developed by Bonvin and co-workers,[11] was used subsequent to HADDOCK. The protocol includes: i) randomization of orientations and rigid body minimization, ii) simulated annealing in torsion angle space, and iii) refine-

Table 2. Residues showing strong CSP and flexible residues.

System	CSPs	Flexible residues
PTP1b–**1**	49, 52, 83 85, 86, 122, 182, 184, 254, 257, 262, 263	24, 45–52, 83–87, 115–124, 181–185, 214–223, 254, 257–266
PKA–**2**	50, 52, 55, 57, 59, 70, 71, 123, 322	49–59, 70–72, 104, 120–129, 170–173, 183–184, 322–327
p38–**3**	33, 35, 38, 42, 53, 57, 80[a], 104, 109, 110, 111, 142[a], 154	29–41, 49–57, 72–76, 83–89, 102–112, 154–159, 166–173

[a] Residue was not included in AIR.

ment in Cartesian space with explicit water. Table 3 summarizes the parameters of each step. CNS[32] was used for all structure calculations. Energies were evaluated by using the full electrostatic and van der Waals energy terms with a cutoff distance of 8.5 Å. Optimized parameters for liquid simulation (OPLS) bonded parameters from a modified version of parallhdg5.2.pro force field were used. Residues located near the binding site were defined as flexible (Table 3). The analysis of the simulations was performed with in house MATLAB-scripts. In the case of SB203580 (**3**), the rotational symmetry of pyridyl- and fluorophenyl rings were accounted for by the calculation of rmsd relative to the reference structure. Initial conformers of ligands were generated manually by using the PyMOL[33] graphics system with a subsequent energy minimization in CNS.

Preparation of the SOS sample and STD measurement: Protein kinase p38 with uniformly deuterated but selectively protonated isoleucine residues (^{1}H-Ile) was prepared as described,[34] with minor modifications. STD-NMR spectra[35] were measured on a Bruker DRX600 spectrometer with 150 ms presaturation (three 50 ms Gaussian-shaped pulses).

[1] H.-J. Böhm, G. Klebe, H. Kubinyi, *Wirkstoffdesign*. Spektrum, Heidelberg **1996**.
[2] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, *Nat. Rev. Drug Discov.* **2004**, *3*, 935–949.
[3] T. L. Blundell, H. Jhoti, C. Abell, *Nat. Rev. Drug Disccov.* **2002**, *1*, 45–54.
[4] G. Otting, K. Wüthrich, *Q. Rev. Biophys.* **1990**, *23*, 39–96.
[5] D. C. Rees, R. Carr, M. Congreve, C. W. Murray, *Nat. Rev. Drug Discov.* **2004**, *3*, 660–672.
[6] M. Pellecchia, D. Meininger, Q. Dong, E. Chang, R. Jack, D. S. Sem, *J. Biomol. NMR* **2002**, *22*, 165–1733
[7] F. Ni, *Prog. Nucl. Magn. Reson. Spectrosc.* **1994**, *26*, 517–606.
[8] a) T. Carlomagno, I. C. Felli, M. Czech, R. Fischer, M. Sprinzl, C. Griesinger, *J. Am. Chem. Soc.* **1999**, *121*, 1945–1948; b) M. J. J. Blommers, W. Stark, C. E. Jones, D. Head, C. E. Owen, W. Jahnke, *J. Am. Chem. Soc.* **1999**; *121*, 1949–1953.
[9] S. B. Shuker, P. J. Hajduk, R. P. Meadows, S. W. Fesik, *Science* **1996**, *274*, 1531–1534.
[10] M. A. McCoy, D. F. Wyss, *J. Am. Chem. Soc.* **2002**, *124*, 11758–11763.
[11] a) C. Dominuez, R. Boelens, A. M. Bonvin, *J. Am. Chem. Soc.* **2003**, *125*, 1731–1737; b) A. D. J. Dijk, R. Boelens, A. M. J. J. Bonvin, *FEBS Lett.* **2005**, *272*, 293–312.
[12] J. P. Linge, S. I. O'Donoghue, M. Nilges, *Meth. Enzymol.* **2001**, *339*, 71–90.
[13] a) M. Nilges, *J. Mol. Biol.* **1995**, *245*, 645–660; b) M. Nilges, S. I. O'Donoghue, *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 107–139.
[14] a) G. M. Clore, M. A. Robien, A. M. Gronenborn, *J. Mol. Biol.* **1993**, *231*, 82–102; b) F. H.-T. Allain, G. Varani, *J. Mol. Biol.* **1997**, *267*, 338–351.
[15] X. Fradera, J. Mestres, *Curr. Top. Med. Chem.* **2004**, *4*, 687–700.
[16] R. D. Taylor, P. J. Jewsbury, J. W. Essex, *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151–166.
[17] S. L. McGovern, B. K. Shoichet, *J. Med. Chem.* **2003**, *46*, 2895–2907.
[18] H. S. Andersen, L. F. Iversen, C. B. Jeppesen, S. Branner, K. Norris, H. B. Rasmussen, K. B. Moller, N. P. H. Moller, *J. Biol. Chem.* **2000**, *275*, 7101–7108.
[19] Y. A. Puius, Y. Zhao, M. Sullivan, D. S. Lawrence, S. C. Almo, Z. Y. Zhang, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 13420–13425.
[20] S. Meier, Y. C. Li, J. Koehn, I. Vlattas, J. Wareing, W. Jahnke, L. P. Wennogle, S. Grzesiek, *J. Biomol. NMR* **2002**, *24*, 165–166.
[21] R. A. Engh, A. Girod, V. Kinzel, R. Huber, D. Bossemeyer, *J. Biol. Chem.* **1996**, *271*, 25157–26164.
[22] J. Zheng, D. R. Knighton, N.-H. Xuong, S. S. Taylor, J. M. Sowadski, L. F. Ten Eyck, *Protein Sci.* **1993**, *2*, 1559–1573.
[23] T. Langer, M. Vogtherr, B. Elshorst, M. Betz, U. Schieborr, K. Saxena, H. Schwalbe, *ChemBioChem* **2004**, *5*, 1508–1516.
[24] Z. Wang, B. J. Canagarajah, J. C. Boehm, S. Kassisa, M. H. Cobb, P. R. Young, S. Abdel-Meguid, J. L. Adams, E. J. Goldsmith, *Structure* **1998**, *6*, 1117–1128.
[25] Z. Wang, P. C. Harkins, R. J. Ulevitch, J. Han, M. H. Cobb, E. J. Goldsmith, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 2327–2332.
[26] M. Vogtherr, K. Saxena, S. Grimme, M. Betz, U. Schieborr, B. Pescatore, T. Langer, H. Schwalbe, *J. Biomol. NMR* **2005**, in press.
[27] P. Akamine, Madhusudan, J. Wu, N. H. Xuong, L. F. Ten Eyck, S. S. Taylor, *J. Mol. Biol.* **2003**, *327*, 159–171.
[28] M. Groves, Z.-J. Yao, P. R. Roller, T. R. Burke, D. Barford, *Biochemistry* **1998**, *37*, 17773–17783.
[29] a) T. Kortemme, D. Baker, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14116–14121; b) I. Halperin, B. Ma, H. Wolfson, R. Nussinov, *Protein* **2002**, *47*, 409–443.
[30] V. Jayalakshmi, N. Rama Krishna, *J. Magn. Reson.* **2004**, *168*, 36–345
[31] B. Wang, K. Raha, K. M. Merz, Jr., *J. Am. Chem. Soc.* **2004**, *126*, 11430–11431.
[32] A. T. Brunger, P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. Grosse-Kunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson, G. L. Warren, *Acta Crystallogr. D Biol. Crystallogr.* **1998**, *54*, 905–921.
[33] The PyMOL Molecular Graphics System, **2002**, DeLano Scientific, San Carlos, CA, USA.
[34] P. J. Hajduk, J. C. Mack, E. T. Olejniczak, C. Park, P. J. Dandliker, B. A. Beutel, *J. Am. Chem. Soc.* **2004**, *126*, 2390–2398.
[35] M. Mayer, B. Meyer, *Angew. Chem.* **1999**, *111*, 1902–1906; *Angew. Chem. Int. Ed.* **1999**, *38*, 1784–1788.

Table 3. Docking parameters.

Runs	Space	Steps	T [K]	Time step [fs]	Protein backbone	Protein side chains	Ligand
~300	–	minimization	–	–	rigid	rigid	rigid
100	torsion angle	1000	2000→50	8	rigid	rigid	flexible
100	torsion angle	4000	2000→50	8	rigid	flexible	flexible
100	torsion angle	1000	500 → 50	2	flexible	flexible	flexible
40	cartesian	500/500/500	100/200/300	2	$K^{pos} = 5$ kcal mol^{-1} Å^{-2}	flexible	$K^{pos} = 5$ kcal mol^{-1} Å^{-2}
40	cartesian	5000/1000/1000	300/200/100	2	flexible	flexible	flexible